日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application:	２００３年　７月２４日
出願番号 Application Number:	特願２００３－２００８４０
[ST. 10/C]:	[ＪＰ２００３－２００８４０]
出願人 Applicant(s):	株式会社日立製作所

U.S. Appln Filed 9-18-03
Inventor: K Shimada
mattingly Stanger & malur
Docket H-1112

２００３年　８月１９日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫

出証番号　出証特２００３－３０６７５２２

【書類名】　特許願
【整理番号】　K03008761A
【あて先】　特許庁長官殿
【国際特許分類】　G06F　3/06
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
　【氏名】　島田　健太郎
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社　日立製作所
【代理人】
　【識別番号】　100075096
　【弁理士】
　【氏名又は名称】　作田　康夫
【手数料の表示】
　【予納台帳番号】　013088
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1
【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　ストレージシステム及びストレージシステムの書込みを高速化する方法

【特許請求の範囲】

【請求項１】

記憶装置、及び

前記記憶装置及び上位装置と接続される第一及び第二の制御部とを有し、

前記第一の制御部は、

第一のメモリ及び第二のメモリを有し、

前記第二の制御部は、

第三のメモリを有し、

前記第一の制御部が前記上位装置からデータを受信した場合は、前記第一の制御部は、前記第一及び第二のメモリに前記データを各々格納して前記上位装置に応答を送信し、

その後、前記第一の制御部は、前記第二のメモリに格納されたデータを前記第三のメモリに転送することを特徴とするストレージシステム。

【請求項２】

更に第一の電源及び第二の電源を有し、

前記第一の制御部のうち、前記第一のメモリは前記第一の電源から電源供給を受け、

前記第一の制御部の前記第二のメモリ及び前記第二の制御部の前記第三のメモリは前記第二の電源から電源供給を受けることを特徴とする請求項１記載のストレージシステム。

【請求項３】

更に第一、第二及び第三の電源を有し、

前記第一のメモリは前記第一の電源から電源供給を受け、

前記第二のメモリは前記第二の電源から電源供給を受け、

前記第三のメモリは前記第三の電源から電源供給をうけることを特徴とする請求項１記載のストレージシステム。

【請求項４】

更に電源を有し、

前記第一のメモリ及び前記第二のメモリは前記電源から電源供給を受け、

前記第二のメモリは更に充電池を有し、前記第二のメモリは前記電源を用いて前記充電池を充電することを特徴とする請求項１記載のストレージシステム。

【請求項５】

前記電源に障害が発生した場合、前記第二のメモリは、前記電源からの電源供給を前記充電力の電源供給に切り替える手段を有することを特徴とする請求項４記載のストレージシステム。

【請求項６】

前記第二のメモリはFIFOバッファであることを特徴とする請求項２記載のストレージシステム。

【請求項７】

前記記憶装置は複数あることを特徴とする請求項６記載のストレージシステム。

【請求項８】

前記第二のメモリは、当該第二のメモリに格納されているデータの有無を示す手段を有することを特徴とする請求項７記載のストレージシステム。

【請求項９】

上位装置と接続されるホストインターフェース部と、

前記ホストインターフェースと接続されるスイッチ部と

前記スイッチ部と接続される第一及び第二の制御部と、

前記第一及び第二の制御部と接続される記憶装置とを有し、

前記第一の制御部は、

第一のメモリ及び第二のメモリを有し、

前記第二の制御部は、

第三のメモリを有し、

前記第一の制御部が前記上位装置からデータを受信した場合は、前記第一の制御部は、前記第一及び第二のメモリに前記データを各々格納して前記上位装置に

応答を送信し、

その後、前記第一の制御部は、前記第二のメモリに格納されたデータを前記第三のメモリに転送することを特徴とするストレージシステム。

【請求項１０】

ディスクドライブと、

上位装置から前記ディスクドライブへ書き込むデータを一時的に格納する第一のキャッシュ手段と、

前記ディスクドライブへ書き込むデータの複製を格納する第二のキャッシュ手段と、

前記第二のキャッシュ手段へ上位装置からの書き込みデータの複製を送るために一時的に格納するFIFOバッファ手段有し、

前記第一のキャッシュ手段に前記上位装置からの書き込みデータを格納し、前記FIFOバッファ手段に前記上位装置からの書き込みデータの複製が格納された時点において上位装置にデータ書き込みの完了の通知を行うことを特徴とするストレージシステム

【請求項１１】

前記第一のキャッシュ手段に接続する第一の電源と、前記第一の電源と独立した、前記第二のキャッシュ手段に接続する第二の電源を備え、前記FIFOバッファ手段は前記第二の電源に接続することを特徴とする請求項１０記載のストレージシステム

【請求項１２】

前記第一のキャッシュ手段に接続する第一の電源と、前記第一の電源と独立した、前記第二のキャッシュ手段に接続する第二の電源と、前記第一の電源と独立した、前記FIFOバッファ手段に接続する第三の電源を備えることを特徴とする請求項１０記載のストレージシステム

【請求項１３】

前記第一のキャッシュ手段に接続する第一の電源と、前記第一の電源と独立した、前記第二のキャッシュ手段に接続する第二の電源を備え、前記FIFOバッファ手段が前記第一の電源に接続し、かつ前記第一の電源が障害を生じた時に前記FI

FOバッファに前記第一の電源に代わって電力を供給するバッテリ手段を備えること特徴とする請求項１０記載のストレージシステム。

【請求項１４】

前記FIFOバッファ手段から前記第二のキャッシュ手段に前記上位装置からの書き込みデータの複製をすべて書き込んだかどうかを表示する残データ有無表示手段を備えることを特徴とする請求項１１記載のストレージシステム。

【請求項１５】

二重化キャッシュを有するストレージシステムにおいてデータを書き込む方法であって、

上位装置からの書き込みデータを２重化して蓄える２重化キャッシュメモリの一方にデータを書き込むステップと、

前記２重化キャッシュメモリの他方よりも高速に書き込みを行えるFIFOバッファにデータを書き込むステップと、

前記２重化キャッシュメモリの一方とFIFOバッファにデータが正しく書き込めたことを確認して上位装置にデータ書き込みの完了を報告するステップと、

上位装置にデータ書き込みの完了を報告した後で前記FIFOバッファに書き込んだデータを前記２重化キャッシュメモリの他方に書き込んでデータの２重化を完了するステップを有することを特徴とする方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、ディスクドライブ等の記憶装置を複数備えるストレージシステムに係わり、特に、計算機等の上位装置からのデータの書き込みを高速化する制御技術に関する。

【０００２】

【従来の技術】

多数の記憶装置を搭載するストレージシステムでは、従来から半導体メモリ等のキャッシュメモリを用いて上位装置と記憶装置との間で授受されるデータを一時的に格納することで、データの読み出し及び書き込みを高速化することが行わ

れてきた。例えば、上位装置からのデータの書き込みの際は、ストレージシステムは書き込みデータを一旦キャッシュメモリに格納し、その時点において上位装置に書き込み処理完了を通知する。その後、ストレージシステムは、上位装置の動作とは独立に、キャッシュメモリに格納したデータを実際に記憶装置に書き込む。

【０００３】

このように、キャッシュメモリを用いてストレージシステムのデータの授受の高速化を行う場合、ストレージシステムは、キャッシュメモリに一旦データを書き込むと記憶装置にデータを書き込むことなく上位装置に書き込み完了を報告してしまう。従って、キャッシュメモリの障害等で記憶装置に書き込む前にキャッシュメモリ内のデータが消失すると、データの回復は非常に困難である。このため、データの消失の可能性を低減しストレージシステムの信頼性を向上させる目的で、ストレージシステムが有するキャッシュメモリ（又はそれを含んだ制御部）を二重化し、データはキャッシュメモリの両方に書き込むという技術が特許文献１及び２に開示されている。

【０００４】

又、キャッシュメモリ又はその冗長化されたメモリを不揮発にすることで、データの損失を防止する技術が、非特許文献１　88頁～89頁に開示されている。

【０００５】

【特許文献１】

特開2001-318766号公報

【特許文献２】

特開平9－146842号公報

【非特許文献１】

「IBM TotalStorage Enterprise Storage Server Model 800」 IBM Redbooks、SG24-6424-01、 Second Edition (October 2002)、IBM Corp.、 ISBN 0738428256

【０００６】

【発明が解決しようとする課題】

特許文献１及び２に開示されたストレージシステムでは、２重化されたキャッシュメモリの双方へのデータの書き込み完了を待って上位装置に完了報告を行うので、時間がかかり、レスポンス性能の向上には限界がある。具体的には、上位装置との間のデータ転送を行う一方の制御部が、２重化された他系の制御部のキャッシュメモリにデータを転送し、かつ他系の制御部のデータの書き込み完了の報告を受けるため、２重系の一方と他方との間の通信の送受が発生し、上位装置からみたストレージシステムのレスポンス性能を下げる主要因となる。一般に２重系では、一方の系と他方の系とは、障害時に相互に影響が出るのを避けるため、電源系を分離するなど、独立性を高めている。このため、一方の系と他方の系の間の通信は時間がかかり、これを高速化することは難しい。

【０００７】

非特許文献１では、一方の制御部内にキャッシュメモリ用のバックアップメモリとして不揮発性メモリを設けているので、上述のような問題、即ち二重系間でのデータ転送による遅延は発生しない。しかし、一般にバッテリで電源を確保している不揮発性メモリでは、不揮発性メモリの容量を大きくすると、バッテリも大容量のものが必要となるので、不揮発メモリの記憶容量を、キャッシュメモリのような大容量にはできない。

【０００８】

このため、キャッシュメモリに書き込むデータの量を不揮発性メモリに格納できる大きさに制限する必要があり、やはり上位装置から見たストレージシステムの書き込み性能を充分高めることは困難である。即ち、書き込むデータの量が不揮発性メモリに格納できる大きさを超えると、キャッシュメモリに書き込んだデータを記憶装置に書き込んでしまうまでは、ストレージシステムは次の上位装置からの書き込みを受け付けることができず、大幅に性能が低下する。

【０００９】

本発明の目的は、キャッシュメモリを用いるストレージシステムにおいて、信頼性向上のためにキャッシュメモリを２重化したときに、上位装置からのデータの書き込みを高速化することである。

【００１０】

【課題を解決するための手段】

本発明では、上記の課題を解決するために、以下の構成を有する。すなわち、ストレージシステムにおいて、複数の制御部及び記憶装置を有する構成とする。ここで、制御部は第一のメモリ及び第二のメモリを有する。第二のメモリは、FIFOバッファ等の第一のメモリよりも容量が少ないメモリでも良い。上記構成において、上位装置からデータの書き込み要求を受けたストレージシステムの制御部は、自己が有する第一のメモリ及び第二のメモリに、書き込み要求に対応するデータを格納する。この時点で、上位装置にデータの書き込み完了を通知する。その後、この制御部は、他の制御部の第一のメモリへ、第二のメモリに格納されたデータを転送する。

【００１１】

本発明では、更に、上記構成において複数の電源を有し、個々の電源が個々の制御部に電源を供給する。ただし、個々の制御部が有する第二のメモリは、その第二のメモリを有する制御部に電源供給している電源とは別の電源から電源供給を受ける構成とする。

【００１２】

尚、第二のメモリも制御部と同一の電源から電源供給を受けるが、第二のメモリが充電池を有し、万一電源が遮断された際には、電源供給元を充電池に切り替える構成としても良い。

【００１３】

更に、複数の制御部と上位装置とを接続するために、ストレージ装置がスイッチ及びインターフェース部を有する構成としても良い。

【００１４】

【発明の実施の形態】

以下、本発明の実施の形態を図面を用いて説明する。

図１は、本発明を適用したストレージシステムの第一の実施形態を示す図である。ストレージシステムは、二つの制御部（以下「コントローラ」）10、二つの電源11及び複数の記憶装置（以下「ディスクドライブ」）12を有する。ここでディスクドライブ12とは、ハードディスクや光ディスク等の記憶媒体を用いた装置

である。一つのディスクドライブ12は、各々双方のコントローラ12と接続されている。又、電源11a及び電源11bは各々独立な電源であり、電源11aはコントローラ10aに電源を供給し、電源11bはコントローラ10bに電源を供給する。したがって、一方の電源11が故障したとしても、他方の電源11がコントローラ10に電源を供給することで、ストレージシステムが動作することが出来る。

【００１５】

各々のコントローラ10は、ホストインターフェース部100、FIFOバッファ102、データ書き込み完了監視部103及びキャッシュメモリ101を有する。尚、ホストインターフェース部やデータ書込み完了監視部は、汎用の演算装置とソフトウエアの組み合わせで実現されても良いし，専用のハードウェアで実現されても良い。

【００１６】

本実施形態のストレージシステムは２重系を構成している。キャッシュメモリ101は、ディスクドライブ12と上位装置との間で授受されるデータを一時的に格納する揮発性の記憶媒体である。

【００１７】

尚、コントローラ10a内のホストインターフェース部100a、キャッシュメモリ101a及びデータ書き込み完了監視部103aは電源線120aを介して電源11aと接続される。又、コントローラ10b内のホストインターフェース部100b、キャッシュメモリ101b及びデータ書き込み完了監視部103bは、電源線120bを介して電源11bと接続される。

【００１８】

このとき、コントローラ10a内のFIFOバッファ102aは、電源線121bを介して電源11bと接続される。更に、コントローラ10b内のFIFOバッファ102bは、電源線121aを介して電源11aと接続される。このように電源線を接続することにより、コントローラ10aのFIFOバッファ102aは、コントローラ10a内の他の構成部分の電源とは異なる、独立した電源に接続されるので、FIFOバッファ102aは、コントローラ10a内のキャッシュメモリ101aと２重系をなす、即ち、一方の電源11が故障した場合にも、他方の電源11から電源供給を受けている部分（キャッシュメモリ101又はFIFOバッファ102）はデータを保持することができる。同様に、コントロー

ラ10b内のFIFOバッファ102bは、コントローラ10b内のキャッシュメモリ101bと２重系をなすことができる。

【００１９】

以下、図１におけるデータ書き込みの手順について簡単に説明する。例えばコントローラ10aに接続される上位装置からデータの書き込みが行われた場合、まずデータをホストインターフェース部100aが受信する。尚、コントローラ10bがデータを受信しても以下の説明と同様の処理が行われる。

【００２０】

ホストインターフェース部100aが受信したデータは、信号線110aを介してFIFOバッファ102a及びキャッシュメモリ101aに送られ、それぞれに書き込まれる。FIFOバッファ102aは、データの書き込みが障害なく完了すると、信号線111aを介して、書き込みの完了をデータ書き込み完了監視部103aに通知する。

【００２１】

キャッシュメモリ101aは、データの書き込みが障害なく完了すると、信号線112aを介して、書き込みの完了をデータ書き込み完了監視部103aに通知する。データ書き込み完了監視部103aは、信号線111a及び信号線112aの両方から書き込み完了の通知を受信したら、信号線113aを介してホストインターフェース部100aに書き込み完了を通知する。通知を受信したホストインターフェース部100aは、上位装置へ書き込み完了を報告する。このとき、コントローラ10aとコントローラ10bの間では特に通信を行わないので、上位装置から見たストレージシステムの書き込み処理を高速化することができる。

【００２２】

FIFOバッファ102aに格納されたデータは、ホストインターフェース部100aより上位装置へ書き込み完了が報告された後、信号線114aを介してコントローラ10b内のキャッシュメモリ101bへ転送され、キャッシュメモリ101bに書き込まれる。このように、コントローラ10aとコントローラ10bとの間のデータ転送に関する通信は、上位装置への書き込み完了が報告された後に行われる。また、FIFOバッファ102aに格納されたデータはキャッシュメモリ101bに移されるので、FIFOバッファ102aを上位装置からの次のデータ書き込みのために空けることができる。この

ため、FIFOバッファ102aの容量が不足してディスクドライブ12にデータを書き込むまで上位装置からの次の書き込みを受け付けられなくなるようなことはなく、さらに書き込み処理の性能を高めることができる。

【００２３】

図２は、FIFOバッファ102aの構成例を示した図である。尚、FIFOバッファ102bも同様の構成となるが、電源11aと11bとが入れ替わった構成となる。FIFOバッファ102aは、電源A監視部301a、書き込み制御部302a、読み出し制御部303a、FIFOメモリ304a、データ確認部305a及び残データ有無指示器330aを有する。

【００２４】

データ確認部305aは、上位装置から送信されてきた書き込みデータが途中で障害を生じて誤ったデータになっていないかCRC等で確認する。データ確認部311aで誤りがないことが確認されると、その結果は信号線311aを介して書き込み制御部302aへ通知される。書き込み制御部302aは、データ確認部305aから誤りがないことが通知されると、書き込み指示信号を信号線313aを介して出力する。書き込み指示信号を受信したFIFOメモリ304aは、上位装置から転送されたデータを格納する。

【００２５】

転送されたデータの格納を終了したFIFOメモリ304aは、信号線111aを介して書き込み完了をデータ書き込み完了監視部103aへ通知する。又、FIFOメモリ304aは、信号線313aを介して、データ書き込みの完了を書き込み制御部302aに通知する。通知を受けた書き込み制御部302aは、信号線315aを介して、読み出し制御部303aにFIFOメモリ304aにデータが格納されたことを通知する。

【００２６】

信号線315aからデータがFIFOメモリ304aに格納されたことを通知された読み出し制御部303aは、信号線314aを介して読み出し指示信号を出力する。読み出し指示信号を受信したFIFOメモリ304aは、格納されたデータを信号線114aに出力し、コントローラ10b内のキャッシュメモリ101bへ送信する。

【００２７】

電源A監視部301aは、図１では図示しない監視信号線310aを用いて電源11aの状

態を監視する。尚、本例において電源11aの状態を監視するのは、以下の理由による。もし、電源11aの完全な障害で給電停止になればホストインターフェース部100aからの信号が消失するので、FIFOバッファ102aは、ホストインターフェース部100aからの信号の有無のみで、電源11aの障害が発見できる。しかし、電源11aが正規の電圧でない不正な電圧を給電するなどすると、ホストインターフェース部100aからの信号が単に消失するのではなく、誤動作を起こしてなんらかの異常な信号となることが予想される。この場合には、ホストインターフェース部100aからの信号のみを監視していては判定がつかない。そこで電源Ａ監視部301aは、電源11aを直接監視することによってホストインターフェース部100aからの信号が正常と期待できるかどうかを判定する。

【００２８】

電源11aに障害が発生し、ホストインターフェース部100aからデータが正常に伝わってこない状態となると、電源A監視部301aは、信号線312aを介して書き込み抑止信号を書き込み制御部302aへ出力し、誤ったデータがFIFOメモリ304aに格納されるのを抑止する。尚、電源11aに障害が発生しても、FIFOバッファ102a内の各部は電源線320a及び電源線121bを介して電源11bに接続されているので、障害なく動作を継続することができる。

【００２９】

また、FIFOバッファ102aは、FIFOメモリ304a内にまだコントローラ10bのキャッシュメモリ101bへ送出されていないデータがあるかどうかを表示する残データ有無指示器330aを備える。この残データ有無指示器330aによって、ストレージシステムの使用者又は管理者は、コントローラ10aで障害が発生した際にコントローラ10aを交換して障害回復を行う時には、データがコントローラ10bのキャッシュメモリ101bへ送出が完了しているかどうかを確認し、データ転送の完了を待って、コントローラ10aを交換することができる。

【００３０】

尚、本発明の第二の実施形態として、ストレージシステムが第一の実施形態の構成に加えて電源43a及びbを有し、FIFOバッファ402aが電源43bに、FIFOバッファ402bが電源43aに接続されている形態も考えられる。本実施形態では、電源43a

及び電源43bは、それぞれFIFOバッファ102a及び102bに電源を供給するだけで良いので、電源11a及び電源11bに比べ、容量を小さくすることができる。

【００３１】

図３は、本発明を適用したストレージシステムの第三の実施形態を示す図である。第三の実施形態は、第一の実施形態と異なり、ストレージシステムは、バッテリ付きFIFOバッファ502a及び502bを有する。又、第一の実施形態と異なり、本実施形態では、バッテリ付きFIFOバッファ502aは、電源線521bを介して電源11aに、バッテリ付きFIFOバッファ502bは、電源線521aを介して電源11bに接続される。バッテリ付きFIFOバッファ502a及び502bは、それぞれ電源11a、電源11bの障害発生時には、内蔵しているバッテリにより動作を継続する。その他の構成及び動作は第一の実施形態と同様である。

【００３２】

図４は、バッテリ付きFIFOバッファ502aの構成例を示す図である。尚、バッテリ付きFIFOバッファ502bもほぼ同一の構成だが、電源11aが電源11bとなる。バッテリ付きFIFOバッファ502aは、電源A監視部301a、書き込み制御部302a、読み出し制御部303a、FIFOメモリ304a、データ確認部605a、残データ有無指示器330a、バッテリ606a、充電監視部607a及び電源セレクタ608aを有する。

【００３３】

電源A監視部310aは、図３に図示しない監視信号線310aを使用して、電源11aの状態を監視している。電源11aに障害が発生したことを監視信号線310aを通じて検出すると、電源A監視部310aは、第一の実施形態と同様に、信号線312aを介して書き込み抑止信号を書き込み制御部302aへ出力し、FIFOメモリ304aへの新たな書き込みデータの格納を抑止させる。また、本実施形態では、電源A監視部301は、書き込み抑止信号の出力にあわせて、信号線616aを介して電源切り替え信号を電源セレクタ608aに出力する。電源切り替え信号を受信した電源セレクタ608aは、各部への電源線620aに接続する電源を電源11aからバッテリ606aに切り替える。

【００３４】

バッテリ606aは、電源11aが正常である間は、充電監視部607aによって充電さ

れている。これにより、バッテリ606aは、いつでも電源11aに代わって各部に電源を供給する準備が整えられている。電源A監視部301aより電源切り替え信号が出力されると、バッテリ606aは、電源線620を介してFIFOバッファ502a内の各部へ電源を供給する。このバッテリ606aからの電源の供給は、電源11aからバッテリ606aへ電源が切り替わった時点にFIFOメモリ304a内に格納されていたデータがすべてコントローラ50bのキャッシュメモリ101bに送出し終わるまでで良い。何故なら、バッテリ606aに電源が切り替わった後は、電源A監視部301aより書き込み抑止信号が出力されるので、新たなデータのFIFOメモリへの格納が生じないからである。このため、バッテリ606aは大きな容量を必要とせず、ごく小さな容量のバッテリで充分である。

尚、上述以外のバッテリ付きFIFOバッファ502aの動作は、第一の実施形態と同様である。

【００３５】

図５は、図３のコントローラ50の具体的な実装例を示す図である。マザーボード70上に、上位装置からの信号線と接続されるホスト接続用コネクタ71、ホストインターフェース部に対応するLSI72、書き込み完了監視部に相当するLSI74、DIMM(Dual In-line Memory Module)基板で実装されるキャッシュメモリ73及びFIFOバッファ用ドータカード75が装着される。さらに、FIFOバッファ用ドータカード75には、FIFOメモリに相当するLSI 751、書き込み制御部、読み出し制御部、電源A監視部、データ確認部及び充電監視部を含むLSI 752並びにバッテリ753が搭載されている。

【００３６】

マザーボード70は、カードエッジ接続部76を介して、他のコントローラ50、ディスクドライブ12、電源11a及び電源11bに接続される。これは、マザーボード70上の各装置には電源11a又は電源11bにより電源が供給されることを意味する。FIFOバッファ用ドータカード75上の各装置には、電源11a、電源11b又はバッテリ753により電源が供給される。

【００３７】

図５では更に、残データ有無指示器754がLEDでマザーボード70のホスト接続用

コネクタ71が配置された方のカードエッジに並べて実装される。こうすることにより、実際のストレージシステムにコントローラ50を組み込む時には、マザーボード70のホスト接続用コネクタ71のあるカードエッジだけがストレージシステム外部から観察可能なようにしておくことが出来る。このようにすることによって、ストレージシステムの使用者又は管理者が、ストレージシステム外部から観察して、当該コントローラ50が障害時に交換可能になったかどうかを容易に知ることができる。

【００３８】

さらに、残データ有無指示器754をFIFOバッファ用ドータカード75上ではなくてマザーボード70上に搭載するため、マザーボード70に電源を配給する電源11からの電源供給が停止しても残データ有無指示器754にFIFOバッファ用ドータカード75からの信号線を介して電源供給ができるように、残データ有無指示器754とFIFOバッファ用ドータカード75とを配線しておく。

【００３９】

又、万一誤ってFIFOメモリに未だデータが存在するのに残データ無しで交換可能との表示になってしまわないように、例えば残データ有無指示器754をLEDで実装するときは、点灯が交換可能であることを意味し、消灯が交換不可を意味するとしておく。これを逆に点灯が交換不可、消灯が交換可能という意味にすると、使用者又は管理者は、マザーボード70上の障害によりFIFOバッファ用ドータカード75からの信号線がマザーボード70上で断線して残データ有無指示器754が消灯した状態と、交換可能で残データ有無指示器754が消灯した状態とが区別できなくなる。

【００４０】

尚、マザーボード70上の障害により残データ有無指示器754が消灯している状態と交換不可で残データ有無指示器754が消灯している状態との区別は、別の手段で行う事もできる。例えば、FIFOメモリから他方のコントローラ50のキャッシュメモリへすべてのデータを送出するまでにかかる最大時間を予め測定等を行い調べておけば、データを送出する最大時間を超えて残データ有無指示器754が消灯したままであるときは、マザーボード70上の障害と考えることができる。

【００４１】

あるいは、上位装置から最後に書き込まれたデータが何であるかが上位装置において知ることができれば、書き込まれたデータが他方のコントローラのキャッシュメモリまたはディスクドライブに格納されているかどうか、他方のコントローラ50から調べるという方法もある。

【００４２】

図６は、本発明を適用したストレージシステムにおけるデータの書き込み方法の手順を示したフローチャートである。本フローチャートは、上述した全ての実施形態で共通である。

【００４３】

まずストレージシステムのコントローラ10aは、ホストインターフェース部100aを介して、上位装置から書き込みコマンドを受信する（ステップ801）。次にコントローラ10aは、書き込みコマンドで指示されたデータの大きさの分だけキャッシュメモリ101aの空き領域を確保する（ステップ802）。次にコントローラ10aは、FIFOバッファ102aの空きを検査し（ステップ803、804）、空きがない場合には、コントローラ10aは、ステップ803及び804の処理をFIFOバッファ102aが空くまで繰り返す。

【００４４】

FIFOバッファに空きが有る場合又はできた場合には、コントローラ10aは、上位装置に書き込み準備完了を通知する（ステップ805）。その後、コントローラ10aは、実際に上位装置からデータを受信する（ステップ806）。データを受信したコントローラ10aは、ホストインターフェース部100aで受信したデータの複製を行う。ただし、データの複製は、他の部分で行われても良い（ステップ807）。その後、ホストインターフェース部100aは、複製したデータの一方をキャッシュメモリ101aに、他方のデータをFIFOバッファ102aに転送する（ステップ808）。

【００４５】

次にコントローラ10aは、キャッシュメモリ101a及びFIFOバッファ102aのいずれかの書き込みでエラーが発生しなかったかどうか検査する（ステップ809）。

もしいずれかの書き込みでエラーが発生した時には、コントローラ10aは、上位装置に書き込みエラーを報告する（ステップ810）。また両方の書き込みでエラーが生じなかった場合には、コントローラ10aは、上位装置からデータを受信し終わったかどうか検査する（ステップ811）。もし受信し終わっていなければ、コントローラ10aは、全てのデータを受信し終わるまでステップ806～811までの処理を繰り返す。

【００４６】

全データを受信し終わった場合、コントローラ10aは、上位装置に書き込みの完了を報告する（ステップ812）。その後、コントローラ10aは、他系（バックアップ側）のコントローラ10bのキャッシュメモリ101bに空き領域を確保するよう、他系のコントローラ10bに指示を出す。具体的には、例えばFIFOバッファ102aが、書き込み要求信号を直接他系コントローラ10bに出しても良い（ステップ813）。空き領域を確保したコントローラ10aは、FIFOバッファ102aから書き込みデータをバックアップ側のキャッシュメモリ101bに転送して処理を終了する（ステップ814）。

【００４７】

このように処理することにより、ストレージシステムは、データをバックアップ側のキャッシュメモリに格納する前に、上位装置へ書き込み完了を報告することが可能である。

【００４８】

図７は、本発明を適用したストレージシステムの第四の実施形態を示す図である。本実施形態のストレージシステムは、コントローラ90を90a、90b、90c、及び90dの4個有する。それぞれのコントローラ90は、90aと90bが２重系、90cと90dが２重系となっている。即ち、電源11aがコントローラ90a及び90cに、電源11bがコントローラ90b及び90cに電力を供給している。又、本実施形態のストレージシステムは、第三の実施形態の構成に加え、交換スイッチ部905及び独立したホストインターフェース部900を有する。

【００４９】

更に各コントローラ90は、他の実施形態のコントローラが有していたホストイ

ンターフェース部100の代わりに接続スイッチ部904を有する。各ホストインターフェース部900と各コントローラ90の接続スイッチ部904との間は、交換スイッチ部905を介して相互に接続される。尚、ホストインターフェース部900及び交換スイッチ部904は、単数でも複数あっても良い。このようにすることにより、上位装置がホストインターフェース部900a及び900bのどちらに接続されても、4個のコントローラ90a、90b、90c、90dのうち任意のコントローラを利用することができる。

【００５０】

本実施形態では、例えばホストインターフェース部900aに接続される上位装置からデータの書き込みが行われた場合、まず当該データが例えば交換スイッチ部905aを経て、コントローラ90aの接続スイッチ部904aに転送される。接続スイッチ部904aが受信したデータは、信号線110aを介してバッテリ付きFIFOバッファ502aとキャッシュメモリ101aに転送される。

【００５１】

データの書き込みを障害なく完了したバッテリ付きFIFOバッファ502aは、信号線111aを介して、書き込みの完了をデータ書き込み完了監視部103aに通知する。又、データの書き込みを障害なく完了したキャッシュメモリ101aは、信号線112aを介して、書き込みの完了をデータ書き込み完了監視部103aに通知する。データ書き込み完了監視部103aは、信号線111a及び信号線112aの双方を介して書き込み完了の通知を受信したら、信号線113aを介して接続スイッチ部904aに書き込み完了を通知する。接続スイッチ部904aは、交換スイッチ部905a及びホストインターフェース部900aを介して上位装置へ書き込み完了を報告する。このとき、コントローラ90aと他のコントローラ90b、90c、90dの間ではデータ転送のための通信が何ら発生しないので、上位装置から見たストレージシステムの書き込み処理を高速化することができる。

【００５２】

FIFOバッファ502aに格納されたデータは、ホストインターフェース部900aより上位装置へ書き込み完了が報告された後、信号線114aを介してコントローラ90b内のキャッシュメモリ101bへ送信され、キャッシュメモリ101bへ書き込まれる。

このように、コントローラ90aとコントローラ90bとの間の通信は、上位装置へ書き込み完了が報告された後に行われる。また、バッテリ付きFIFOバッファ502aに格納されたデータは、キャッシュメモリ101bに転送されるので、バッテリ付きFIFOバッファ502aを上位装置からの次のデータ書き込みのために空けることができる。このため、バッテリ付きFIFOバッファ502aの容量が不足し、ディスクドライブ12にデータを書き込むまで上位装置からの次の書き込みを受け付けられなくなるようなことはない。以上のようにすることによって、ストレージシステムの書き込み処理の性能を高めることが可能である。

【００５３】

尚、本実施形態のバッテリ付きFIFOバッファ502の動作は、図３の第三の実施形態におけるバッテリ付きFIFOバッファ502と同様である。

【００５４】

【発明の効果】

本発明により、キャッシュメモリを２重化するストレージシステムにおいて、上位装置からみたデータ書き込みの時間を短縮することができる。これにより上位装置からみたデータ書き込み性能を向上させることができる。

【図面の簡単な説明】

【図１】

ストレージシステムの実施の形態の第一の例を示す図である。

【図２】

ストレージシステムに用いられるFIFOバッファの構成の例を示す図である。

【図３】

ストレージシステムの実施の形態の第三の例を示す図である。

【図４】

ストレージシステムに用いられるバッテリ付きFIFOバッファの構成の例を示す図である。

【図５】

ストレージシステムに用いられるコントローラのマザーボードの概観の例を示す図である。

【図６】

ストレージシステムの書き込みを高速化する方法の実施の形態の例を示す図である。

【図７】

ストレージシステムの実施の形態の第四の例を示す図である。

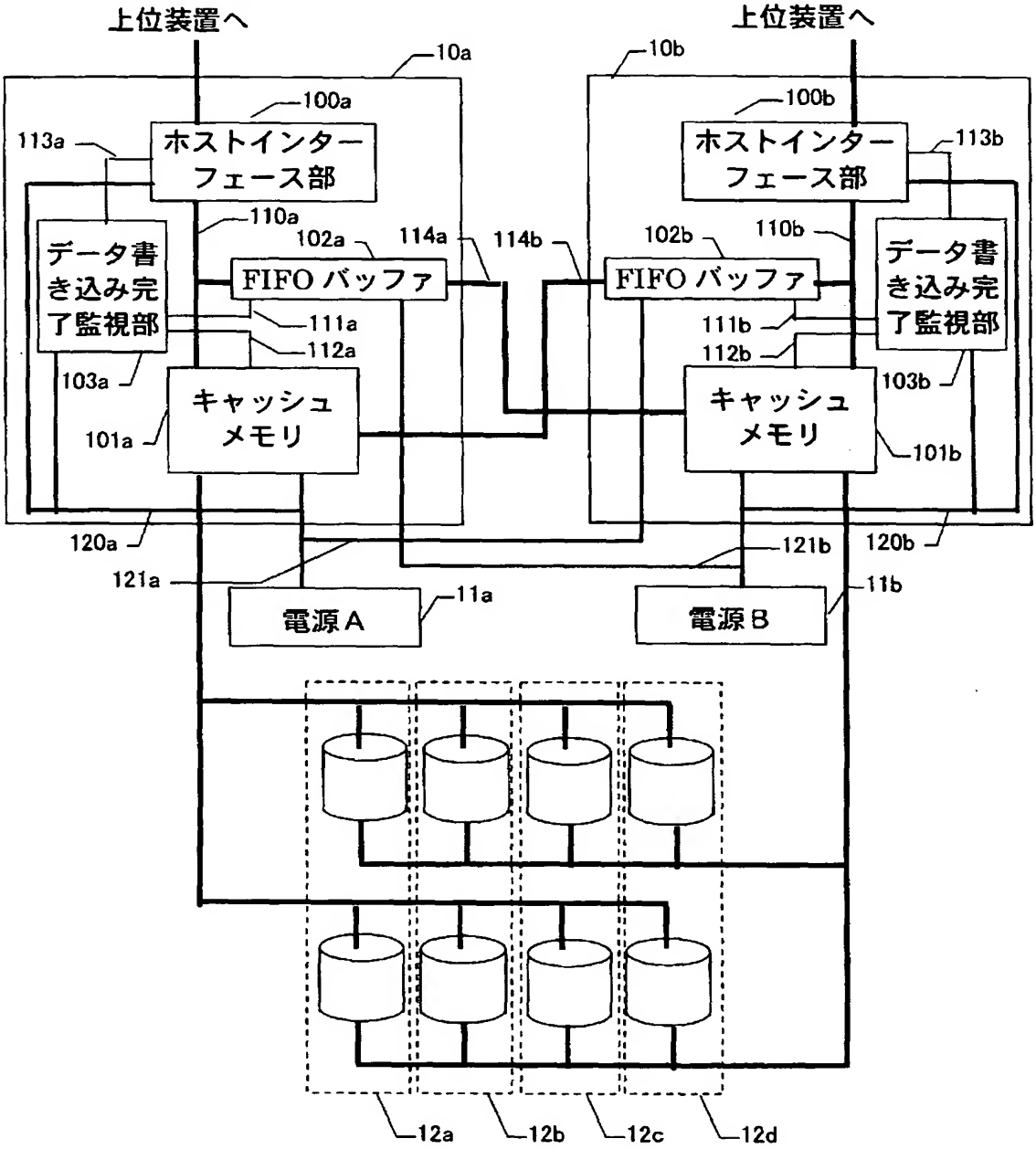
【符号の説明】

10…コントローラ、11a…電源A、11b…電源B、12…ディスクドライブ、43a…小電源A、43b…小電源B、100…ホストインターフェース部、101…キャッシュメモリ、102…FIFOバッファ、502…バッテリ付きFIFOバッファ、103…データ書き込み完了監視部、606…バッテリ、607…充電監視部、608…電源セレクタ。

【書類名】　図面

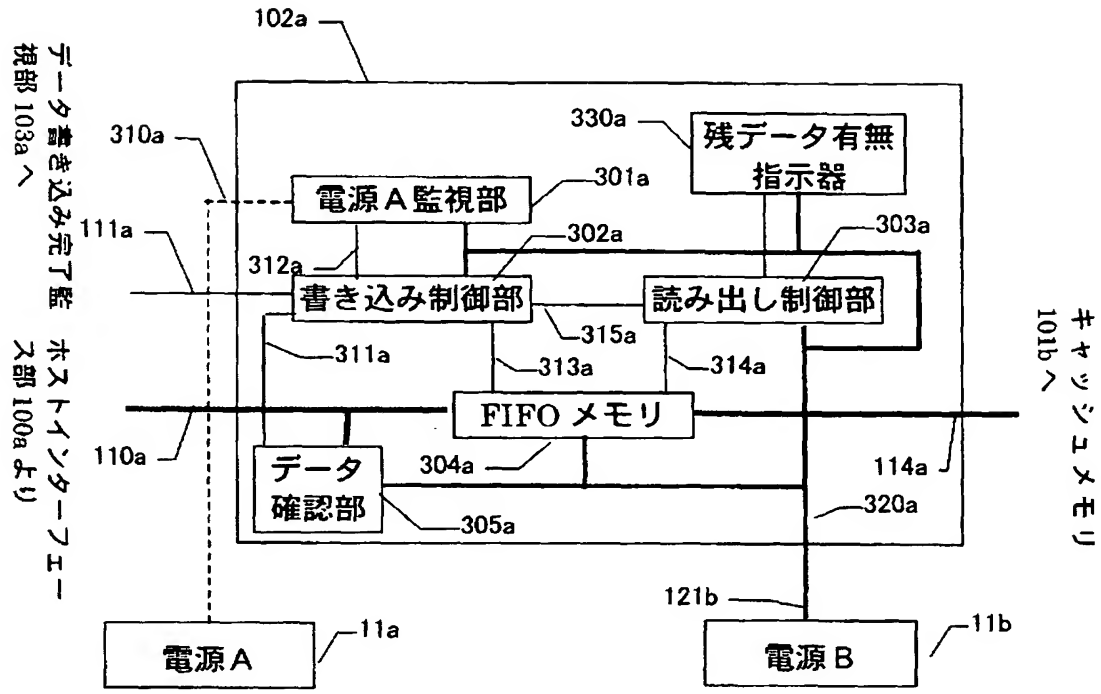
【図１】

図１


上位装置へ
上位装置へ
10a
10b
100a
100b
113a
ホストインターフェース部
ホストインターフェース部
113b
110a
110b
102a
114a
114b
102b
データ書き込み完了監視部
FIFO バッファ
FIFO バッファ
データ書き込み完了監視部
111a
111b
112a
112b
103a
103b
キャッシュメモリ
キャッシュメモリ
101a
101b
120a
120b
121a
121b
電源Ａ
電源Ｂ
11a
11b
12a
12b
12c
12d

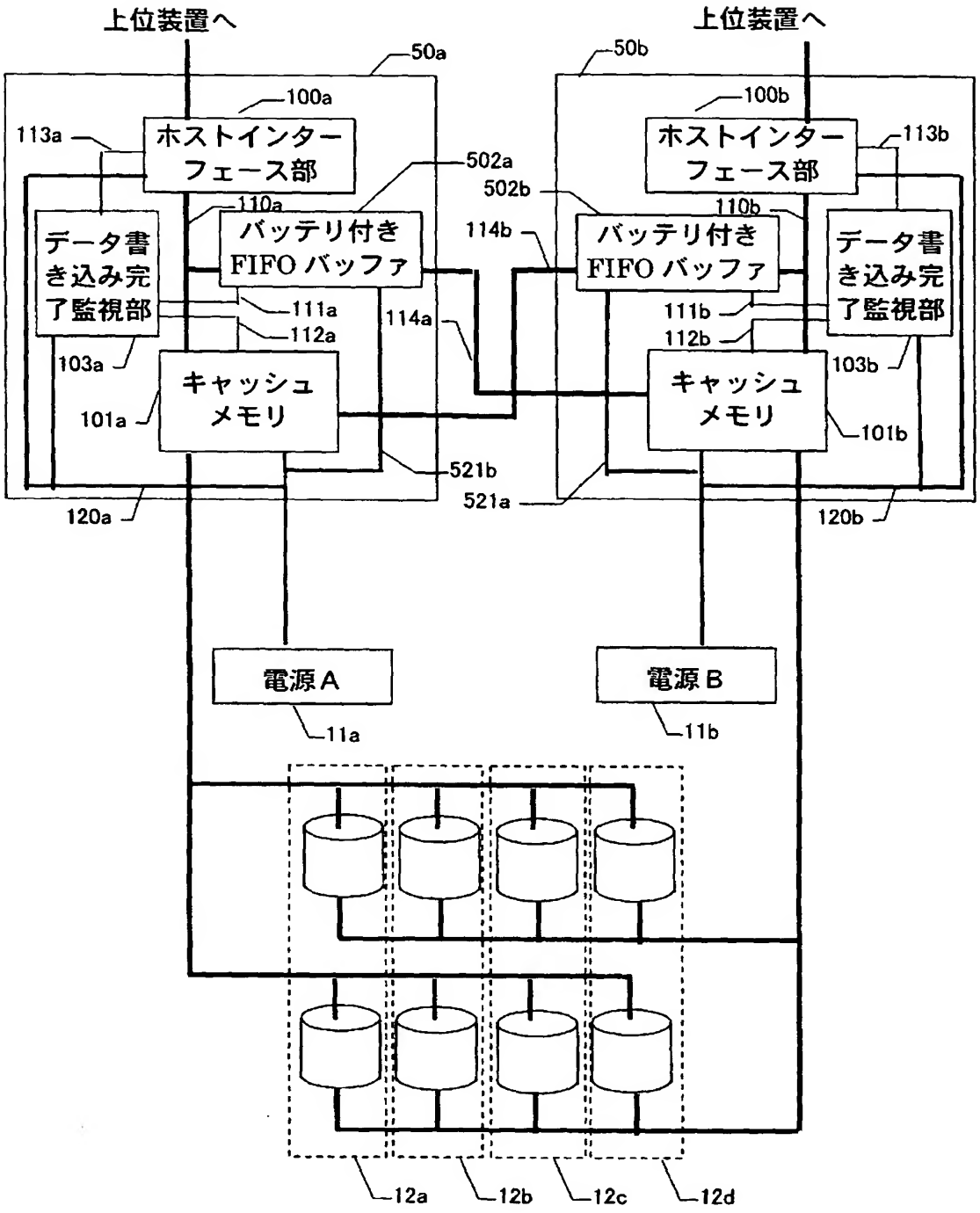
【図２】

図２


102a
データ書き込み完了監視部 103aへ
ホストインターフェース部 100aより
310a
330a
残データ有無指示器
電源Ａ監視部
301a
111a
302a
303a
312a
書き込み制御部
読み出し制御部
315a
311a
313a
314a
キャッシュメモリ 101bへ
FIFO メモリ
110a
データ確認部
304a
114a
305a
320a
121b
11a
11b
電源Ａ
電源Ｂ

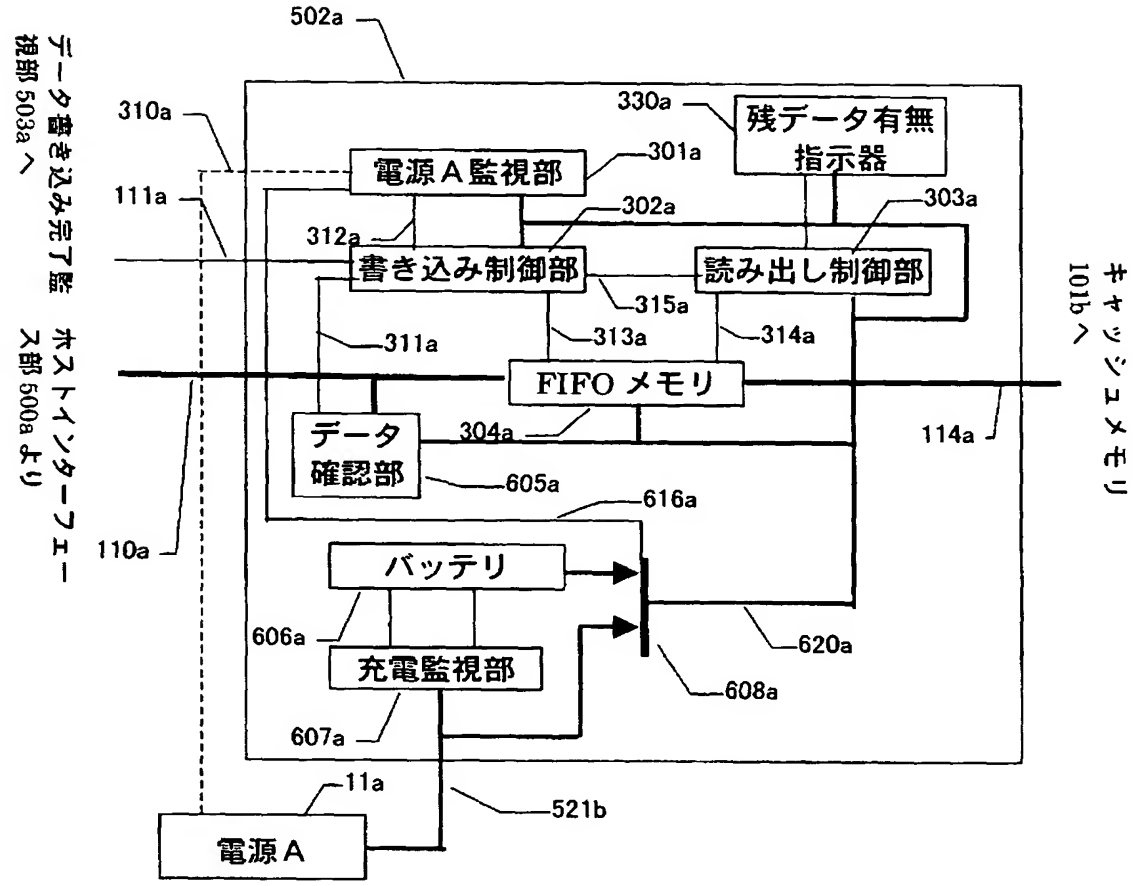
【図３】

図３


上位装置へ
上位装置へ
50a
50b
100a
100b
113a
ホストインターフェース部
ホストインターフェース部
113b
502a
502b
110a
110b
データ書き込み完了監視部
バッテリ付きFIFOバッファ
バッテリ付きFIFOバッファ
データ書き込み完了監視部
114b
111a
112a
114a
111b
112b
103a
103b
101a
キャッシュメモリ
キャッシュメモリ
101b
521b
521a
120a
120b
電源A
電源B
11a
11b
12a
12b
12c
12d

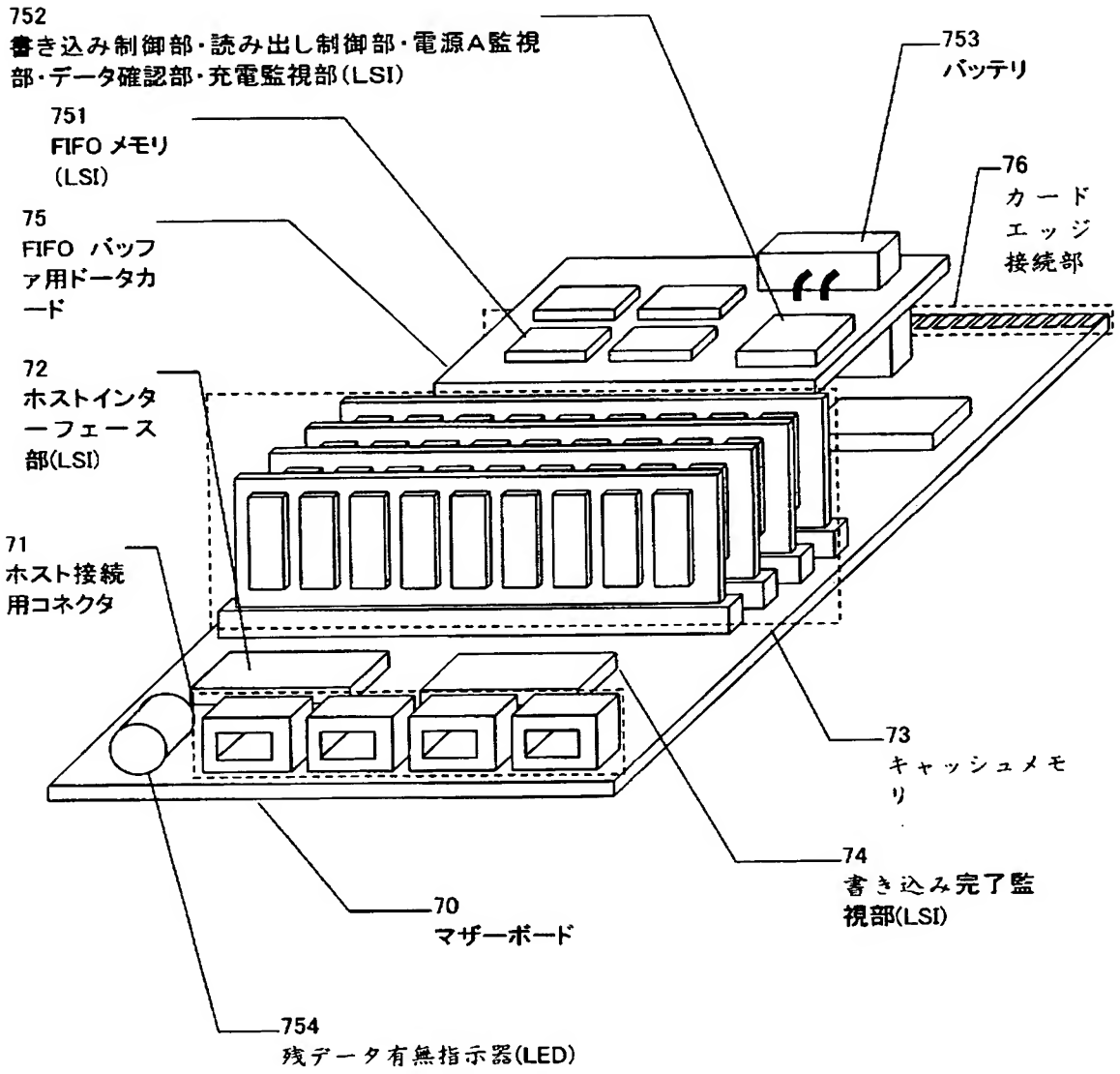
【図４】

図４


502a
データ書き込み完了監視部 503a へ
ホストインターフェース部 500a より
310a
111a
330a
残データ有無指示器
電源Ａ監視部
301a
302a
303a
312a
書き込み制御部
読み出し制御部
315a
311a
313a
314a
FIFO メモリ
キャッシュメモリ 101b へ
データ確認部
304a
114a
605a
616a
110a
バッテリ
606a
620a
充電監視部
608a
607a
11a
521b
電源Ａ

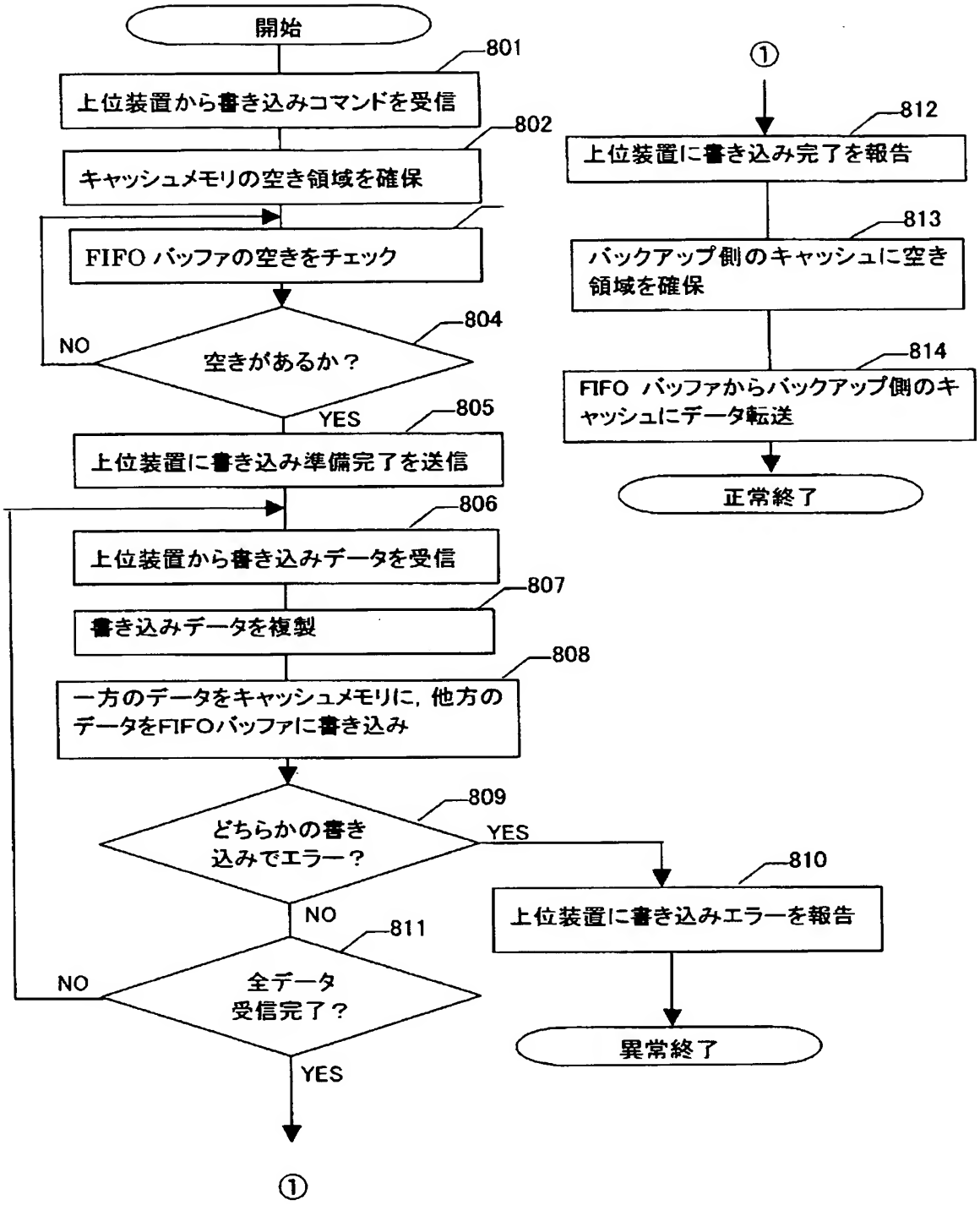
【図５】

図５


752
書き込み制御部・読み出し制御部・電源A監視部・データ確認部・充電監視部(LSI)
753
バッテリ
751
FIFO メモリ
(LSI)
76
カードエッジ接続部
75
FIFO バッファ用ドータカード
72
ホストインターフェース部(LSI)
71
ホスト接続用コネクタ
73
キャッシュメモリ
74
書き込み完了監視部(LSI)
70
マザーボード
754
残データ有無指示器(LED)

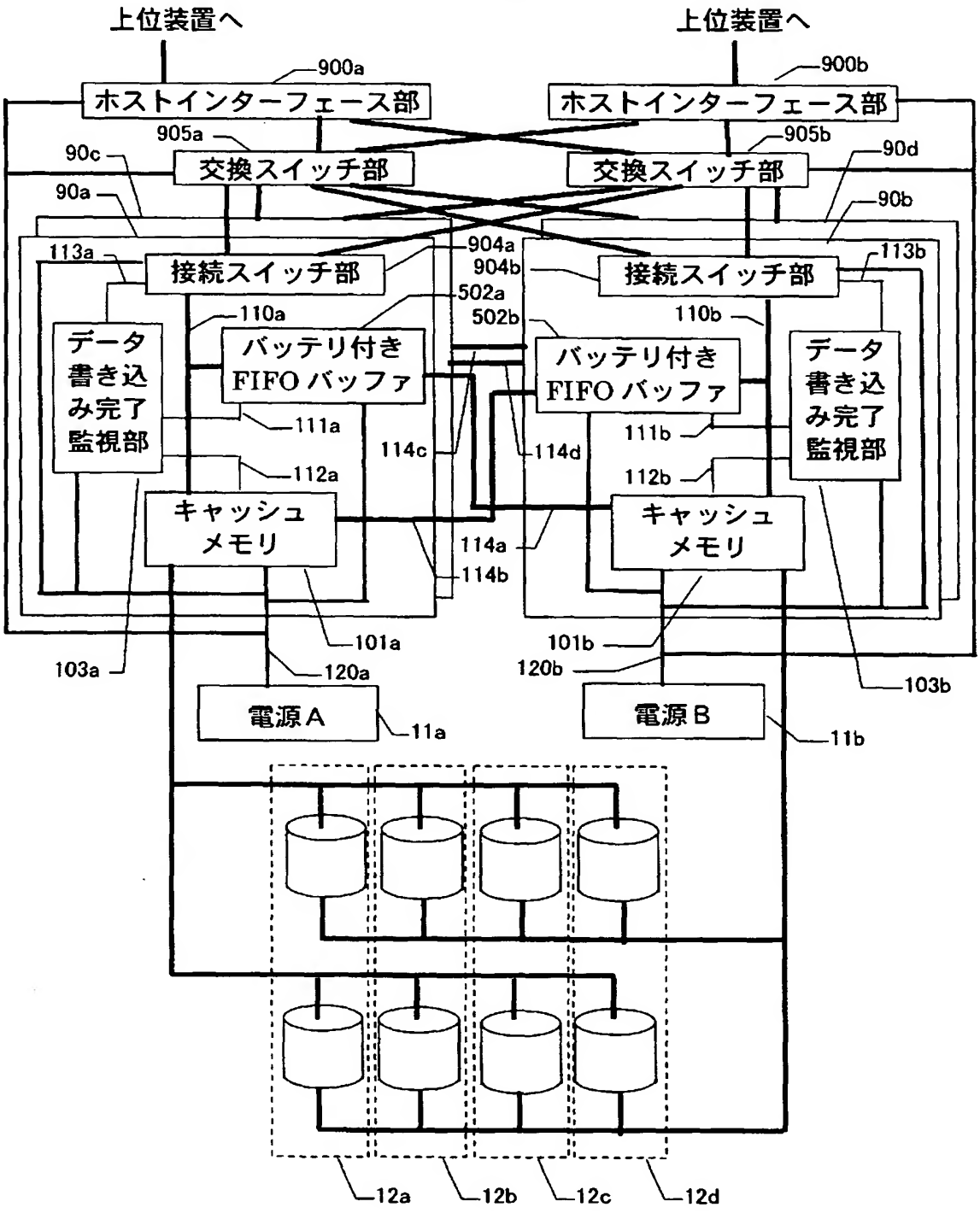
【図６】

図６


開始
801
上位装置から書き込みコマンドを受信
802
キャッシュメモリの空き領域を確保
FIFO バッファの空きをチェック
804
NO
空きがあるか？
YES
805
上位装置に書き込み準備完了を送信
806
上位装置から書き込みデータを受信
807
書き込みデータを複製
808
一方のデータをキャッシュメモリに，他方のデータをFIFOバッファに書き込み
809
どちらかの書き込みでエラー？
YES
NO
811
NO
全データ受信完了？
YES
①
①
812
上位装置に書き込み完了を報告
813
バックアップ側のキャッシュに空き領域を確保
814
FIFO バッファからバックアップ側のキャッシュにデータ転送
正常終了
810
上位装置に書き込みエラーを報告
異常終了

【図７】

図７


上位装置へ
上位装置へ
900a
900b
ホストインターフェース部
ホストインターフェース部
905a
905b
90c
90d
交換スイッチ部
交換スイッチ部
90a
90b
113a
113b
904a
904b
接続スイッチ部
接続スイッチ部
502a
502b
110a
110b
データ書き込み完了監視部
バッテリ付きFIFOバッファ
バッテリ付きFIFOバッファ
データ書き込み完了監視部
111a
111b
114c
114d
112a
112b
キャッシュメモリ
キャッシュメモリ
114a
114b
101a
101b
103a
120a
120b
103b
電源A
電源B
11a
11b
12a
12b
12c
12d

【書類名】　要約書

【要約】

【課題】　キャッシュを２重化するストレージシステムで書き込みを高速化する。

【解決手段】　ディスクドライブ群12、上位装置からディスクドライブ群へ書き込むデータ及びその複製を一時的に格納するキャッシュメモリ101a及び101b、データの複製をキャッシュ101aからキャッシュ101bに送るために一時的に格納するFIFOバッファ102a、データの複製をキャッシュ101bからキャッシュ101aに送るために一時的に格納するFIFOバッファ102bを備える。上位装置からディスクドライブ群へ書き込むデータをキャッシュ101aに格納し、データの複製をキャッシュ101bに格納する前に、キャッシュ101a及びFIFOバッファ102aに書き込みデータ及び複製が格納された時点で書き込みの完了報告を行う。その後、FIFOバッファ102aに格納されたデータの複製をキャッシュ101bに格納する。

【選択図】　図１

認定・付加情報

特許出願の番号	特願２００３－２００８４０
受付番号	５０３０１２１８８３１
書類名	特許願
担当官	第七担当上席　　００９６
作成日	平成１５年　７月２５日

＜認定情報・付加情報＞

【提出日】　平成15年 7月24日

特願２００３－２００８４０

出　願　人　履　歴　情　報

識別番号　　［０００００５１０８］

１．変更年月日　　１９９０年　８月３１日
［変更理由］　　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所